dbGaP Study Release Notes

Release Notes for NHLBI TOPMed WGS VUH AF, phs001032.v6.p2
"NHLBI TOPMed: The Vanderbilt Atrial Fibrillation Registry"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

October	19, 2016	Version 1 Data set release date
January	25, 2017	Version 2 Data set release date
December	13, 2017	Version 3 Data set release date
January	9, 2019	Version 4 Data set release date
January	24, 2020	Version 5 Data set release date
June	3, 2021	Version 6 Data set release date

2021-06-03
Version 6 Data set release for NHLBI TOPMed WGS VUH AF now available

This release includes the addition of Freeze 9 whole genome sequences (WGS) and corresponding VCFs. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (IRB) (GRU-IRB)

Data Type	subjects	samples
Phenotype	1134	1134
Seq_DNA_SNP_CNV (VCFs)	1122	1122
WGS*	1122	1122

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Data Updates

1. **New Study Accession**

NHLBI TOPMed WGS VUH AF version 5 phs001032.v5.p2 has been updated to version 6. The dbGaP accession for the current set of data is **phs001032.v6.p2**. The participant number (p#) has not changed in version 6. No new subjects have been added to this study.

2. There are no updates to the phenotype datasets.

Molecular Data Updates

Two genomic accessions, phg001326.v1 freeze 8 and phg001566.v1 freeze9, are associated with the study.
1. See download components 'sample-info' for manifest of genotyped samples and files.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in 'genotype-qc' tars.

phg	freeze	sample_cnt	subject_cnt
phg001326.v1	8	1110	1110
phg001566.v1	9	1122	1122

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001032/phs001032.v6.p2

2020-01-24

Version 5 Data set release for NHLBI TOPMed WGS VUH AF now available

This release includes the addition of Freeze 8 whole genome sequences (WGS) brokered through the Sequence Read Archive (SRA), and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (IRB) (GRU-IRB)

Data Type	subjects	samples
Phenotype	1134	1134
Seq_DNA_SNP_CNV (VCFs)	1110	1110
WGS*	1110	1110

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Data Updates

1. **New Study Accession**

NHLBI TOPMed WGS VUH AF version 4 phs001032.v4.p2 has been updated to version 5. The dbGaP accession for the current set of data is **phs001032.v5.p2**. The participant number (p#) has not changed in version 5. No new subjects have been added to this study.
2. There are no updated datasets.

Molecular Data Updates

1. See download components phg001326.v1.TOPMed_WGS_Vanderbilt_AF_v5.sample-info.MULTI.tar for manifest of genotyped samples and files.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in phg001326.v1.TOPMed_WGS_Vanderbilt_AF_v5.genotype-qc.MULTI.tar
4. Only Freeze 5b and Freeze 8 VCFs will be available for download.

phg_name	sample_cnt	subject_cnt
phg001138.v1	1018	1018
phg001326.v1	1110	1110

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001032/phs001032.v5.p2

2019-01-09

Version 4 Data set release for NHLBI TOPMed WGS VUH AF now available

This release includes updated phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (IRB) (GRU-IRB)

	Phenotype	Seq_DNA_SNP_CNV (VCFs)	WGS
subjects	1134	1018	1134
samples	1134	1018	1134

Study and Phenotype Data Updates

1. **New Study Accession**

NHLBI TOPMed WGS VUH AF version 3 phs001032.v3.p2 has been updated to Version 4. The

dbGaP accession for the current set of data is **phs001032.v4.p2**. The participant number (p#) has not changed in version 4. No new subjects have been added to this study.

2. **Updated Datasets (n=1)**

pht	version	Dataset Name
5099	3	TOPMed_WGS_VUH_AF_Sample

3. Please note we are discontinuing the submission and distribution of the SAMPLE_USE variable. The sample use counts will be populated by SRA (sequences) and dbGaP (all other submitted molecular data).

Molecular Data Updates

New genomics accession, phg001138 with genotypes on GRC38 is added to the study.
1. See download components phg001138.v1.TOPMed_WGS_Vanderbilt_AF_v4.sample-info.MULTI.tar for manifest of genotyped samples and files.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.

phg_name	sample_cnt	subject_cnt
phg000991.v1	1110	1110
phg001138.v1	1018	1018

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001032/phs001032.v4.p2

2017-12-13

Version 3 Data set release for NHLBI TOPMed WGS VUH AF now available

This release includes updated phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (IRB) (GRU-IRB)

	Phenotype	Seq_DNA_SNP_CNV	Seq_DNA_WholeGenome

subjects	1134	1110	1134
samples	1134	1110	1134

Molecular data descriptions:
(https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi)

a. Seq_DNA_SNP_CNV: SNP and CNV genotypes derived from sequence data (VCFs)
b. Seq_DNA_WholeGenome: Whole genome sequencing

Study and Phenotype Data Updates

1. **New Study Accession**

NHLBI TOPMed WGS VUH AF version 2 phs001032.v2.p1 has been updated to Version 3. The dbGaP accession for the current set of data is **phs001032.v3.p2**. The participant number (p#) has changed in version 3; subjects have been retired. No new subjects have been added to this study.

2. **New Datasets (n=1 dataset)**

pht	version	Dataset Name
7135	1	TOPMed_WGS_VUH_AF_Pedigree

3. **Updated Datasets (n=4 datasets; all existing variables have been updated)**

pht	version	Dataset Name
5098	2	TOPMed_WGS_VUH_AF_Subject
5099	2	TOPMed_WGS_VUH_AF_Sample
5100	2	TOPMed_WGS_VUH_AF_Sample_Attributes
5675	3	TOPMed_WGS_VUH_AF_Subject_Phenotypes

4. **New Variables (n=7 variables)**

pht	pht version	Dataset Name	phv	Variable Name
5098	2	TOPMed_WGS_VUH_AF_Subject	328568	SUBJECT_SOURCE
5098	2	TOPMed_WGS_VUH_AF_Subject	328569	SOURCE_SUBJECT_ID
5100	2	TOPMed_WGS_VUH_AF_Sample_Attributes	328575	SEQUENCING_CENTER
5100	2	TOPMed_WGS_VUH_AF_Sample_Attributes	328576	Funding_Source
5100	2	TOPMed_WGS_VUH_AF_Sample_Attributes	328577	TOPMed_Phase
5100	2	TOPMed_WGS_VUH_AF_Sample_Attributes	328578	TOPMed_Project
5100	2	TOPMed_WGS_VUH_AF_Sample_Attributes	328579	Study_Name

Molecular Data Updates

1. See download components phg000991.v1.TOPMed_WGS_Vanderbilt_AF_v3.sample-info.MULTI.tar and phg000991.v1.TOPMed_WGS_Vanderbilt_AF_v3.marker-info.MULTI.tar for manifest of genotyped samples and files, and information about marker set used for genotyping.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in phg000991.v1.TOPMed_WGS_Vanderbilt_AF_v3.genotype-qc.MULTI.tar.

dbGaP Study Release Notes

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001032/phs001032.v3.p2

2017-01-25

Version 2 Data set release for NHLBI TOPMed WGS VUH AF now available

This release includes changes to the subject phenotype data table. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (IRB) (GRU-IRB)

	phenotype	SRA/VCFs
subjects	1157	310
samples	1157	310

Study and Phenotype Data Updates

1. **New Study Accession**

NHLBI TOPMed WGS VUH AF version 1 phs001032.v1.p1 has been updated to Version 2. The dbGaP accession for the current set of data is **phs001032.v2.p1**. The participant number (p#) has not changed in version 2. No new subjects have been added to this study.

2. **Deleted Variable (n=1)**

pht	Dataset Name	phv	Variable Name
5675	TOPMed_WGS_VUH_AF_Subject_Phenotypes	265820	EF_ever_50

Authorized Access (Individual Level Data and SRA Data)

Individual level data and SRA sequencing data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001032/phs001032.v2.p1

2016-10-19

Version 1 Data set release for NHLBI TOPMed WGS VUH AF now available

This release includes TOPMed Phase I phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Additionally, phenotype tables include subjects and samples beyond TOPMed Phase I in order to instantiate IDs for future versions. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): General Research Use (IRB) (GRU-IRB)

	phenotype	SRA/VCFs
subjects	1157	310
samples	1157	310

Molecular Data Updates

1. See download components phg000796.v1.TOPMed_WGS_Vanderbilt_AF.sample-info.MULTI.tar for manifest of genotyped samples.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in phg000796.v1.TOPMed_WGS_Vanderbilt_AF.genotype-qc.MULTI.tar.

Authorized Access (Individual Level Data and SRA Data)

Individual level data and SRA sequencing data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001032/phs001032.v1.p1